JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Automating Science Operations for Space Missions:
Machine Learning Algorithms for Orbit Region Classification

Second AI and Data Science Workshop for Earth and Space Sciences

February 10, 2021

Kiley Yeakel
Data Scientist
JHU APL SES/SAA
kiley.yeakel@jhuapl.edu

Automated Identification of Regions Around Saturn

Currently scientists spend substantial amounts of time hand-labeling data to identify boundary crossings.

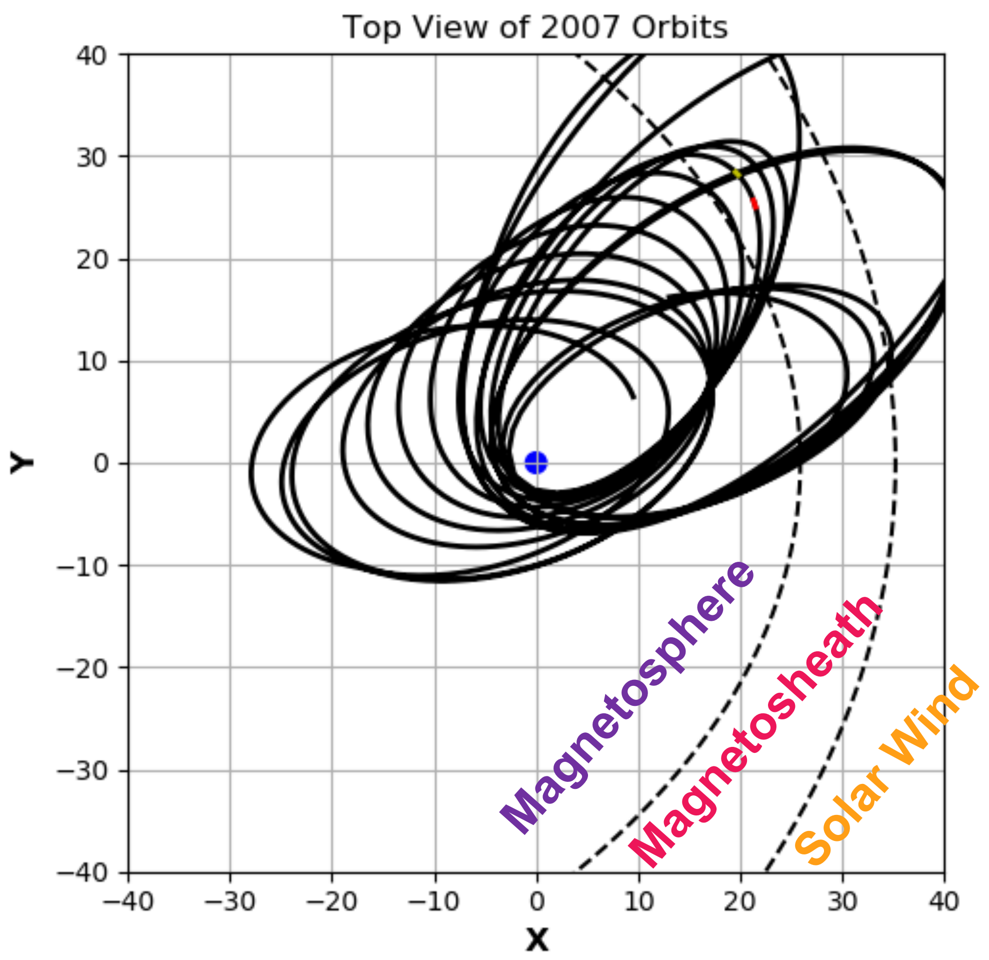
List of ~3k bow shock and magnetopause crossings from the Cassini mission spanning 2004 – 2016 using both magnetometer and CAPS data (until 2012).

Problem 1: Can we automate region selection using more limited datasets (i.e., using only magnetometer data)?

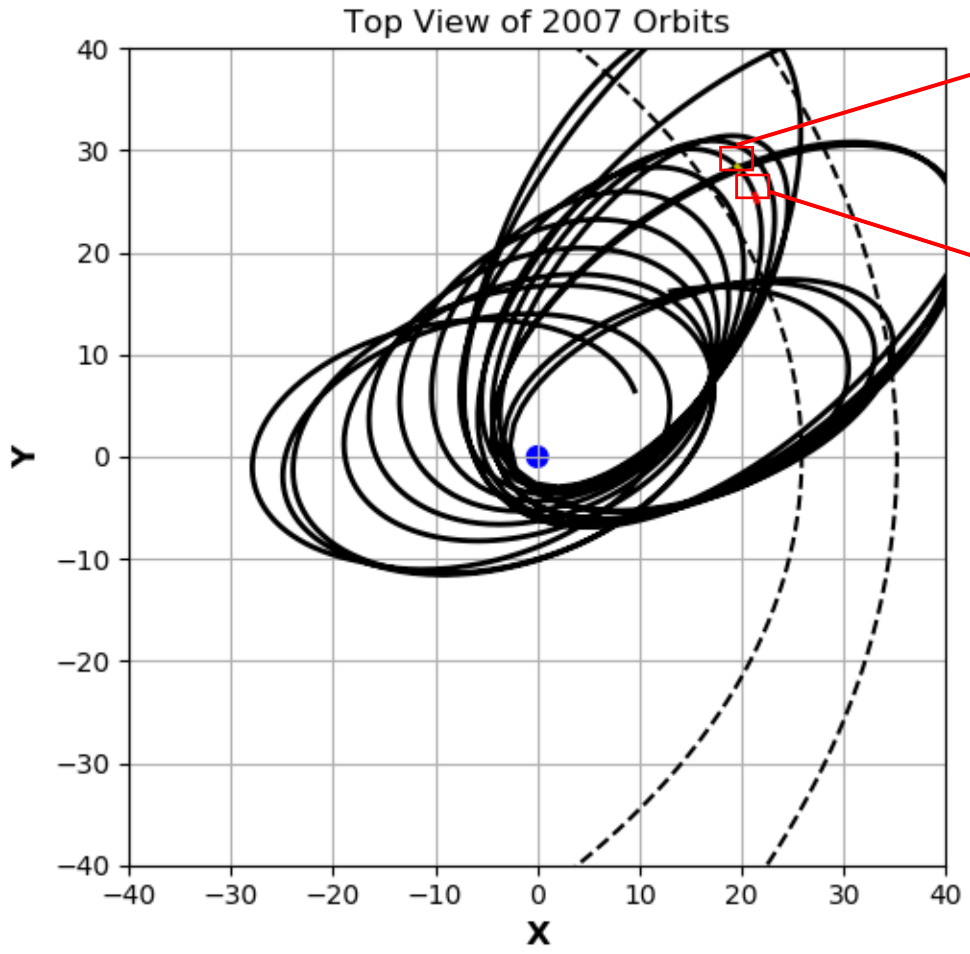
Problem 2: Can we use different datasets (MIMI/CHEMS/LEMMS & MAG & CAPS) and get similar identification results?

Ultimate goal: Develop a proxy algorithm for identification/ classification that will operate on-board the spacecraft

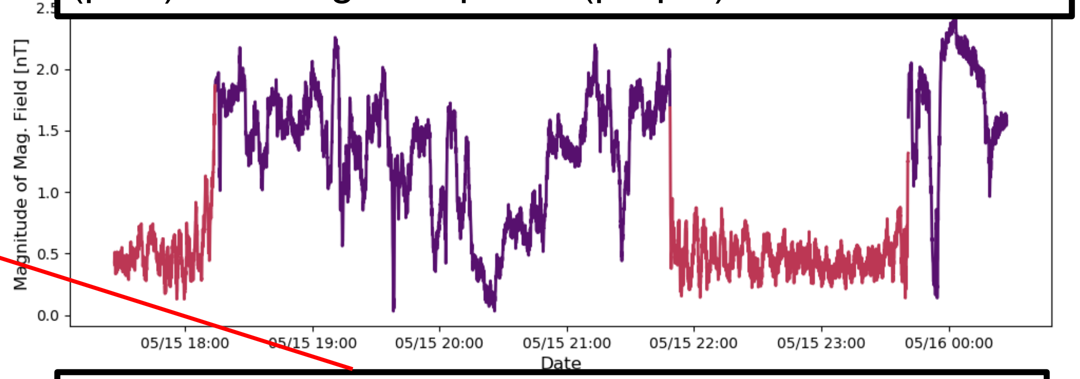
Explored two main approaches: Recurrent Neural Networks (RNN) with feature-limited data, and simpler classifiers including support vector machines (SVM), logistic regression (LR) and random forests (RF)



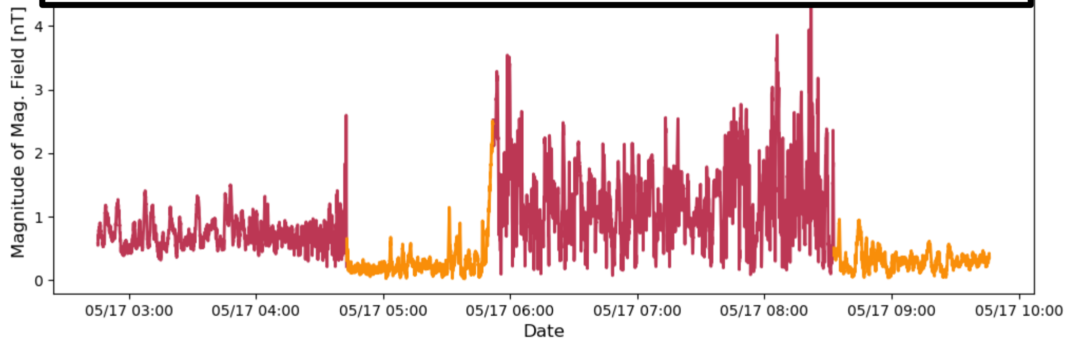
Defining characteristics of the different regions



Example of crossings between the magnetosheath (pink) and magnetosphere (purple)

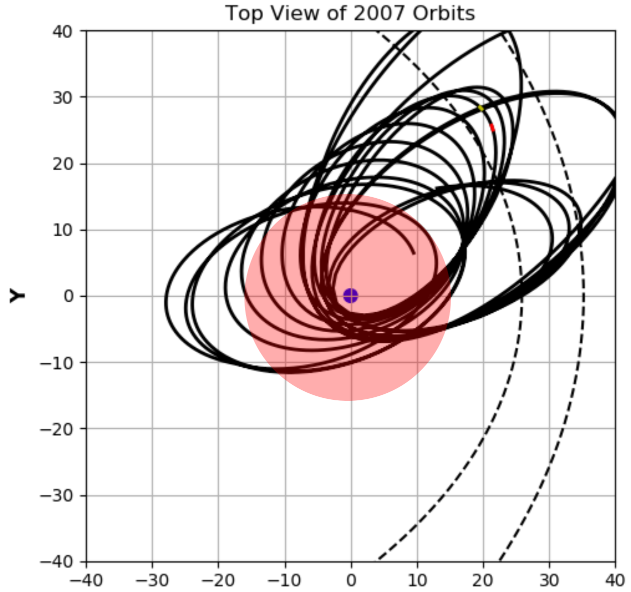


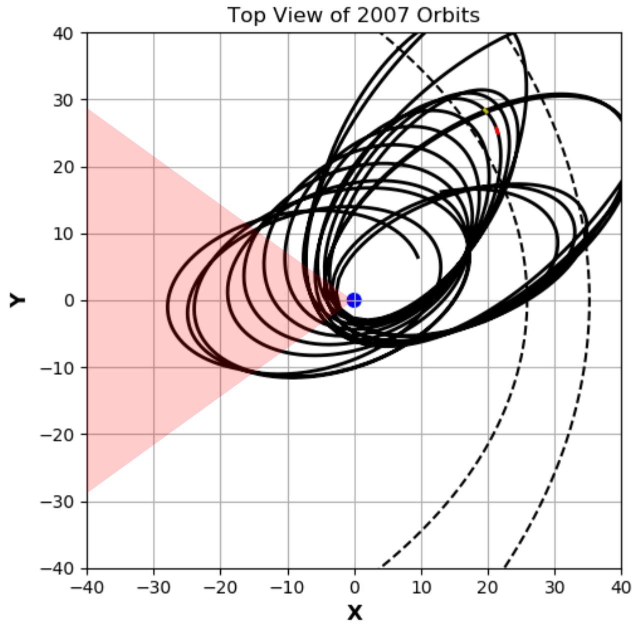
Example of crossings between the magnetosheath (pink) and solar wind (orange)



Clear discrepancy in the running mean and variance of |B| between the different regions – indicates a time series based approach could be useful

Correcting for Imbalanced Datasets





The dataset is heavily weighted towards data collected within the magnetosphere, due to a bias of orbits within that region.

Knowing the geometry of the problem, we can automatically exclude points not close to region transitions with only knowledge of the spacecraft's location.

Low R → within the magnetosphere
Conservatively define an inner radius within which the S/C is definitely within the magnetosphere

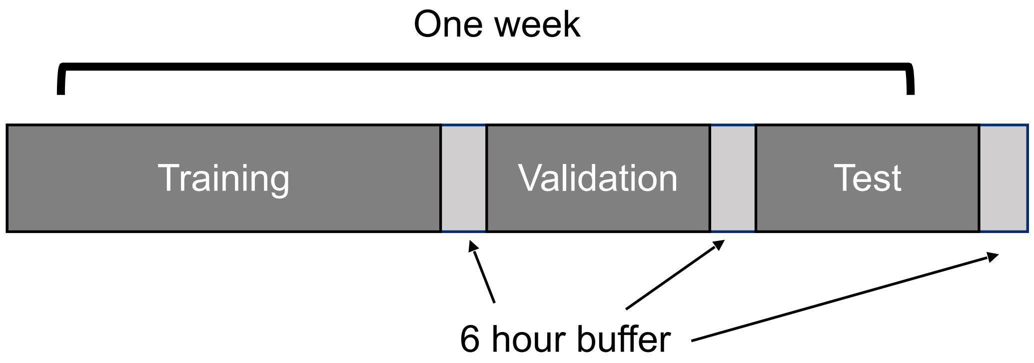
Local time near midnight → within the magnetosphere
Conservatively define a time range for which the S/C is definitely within the magnetosphere

Dataset Preprocessing

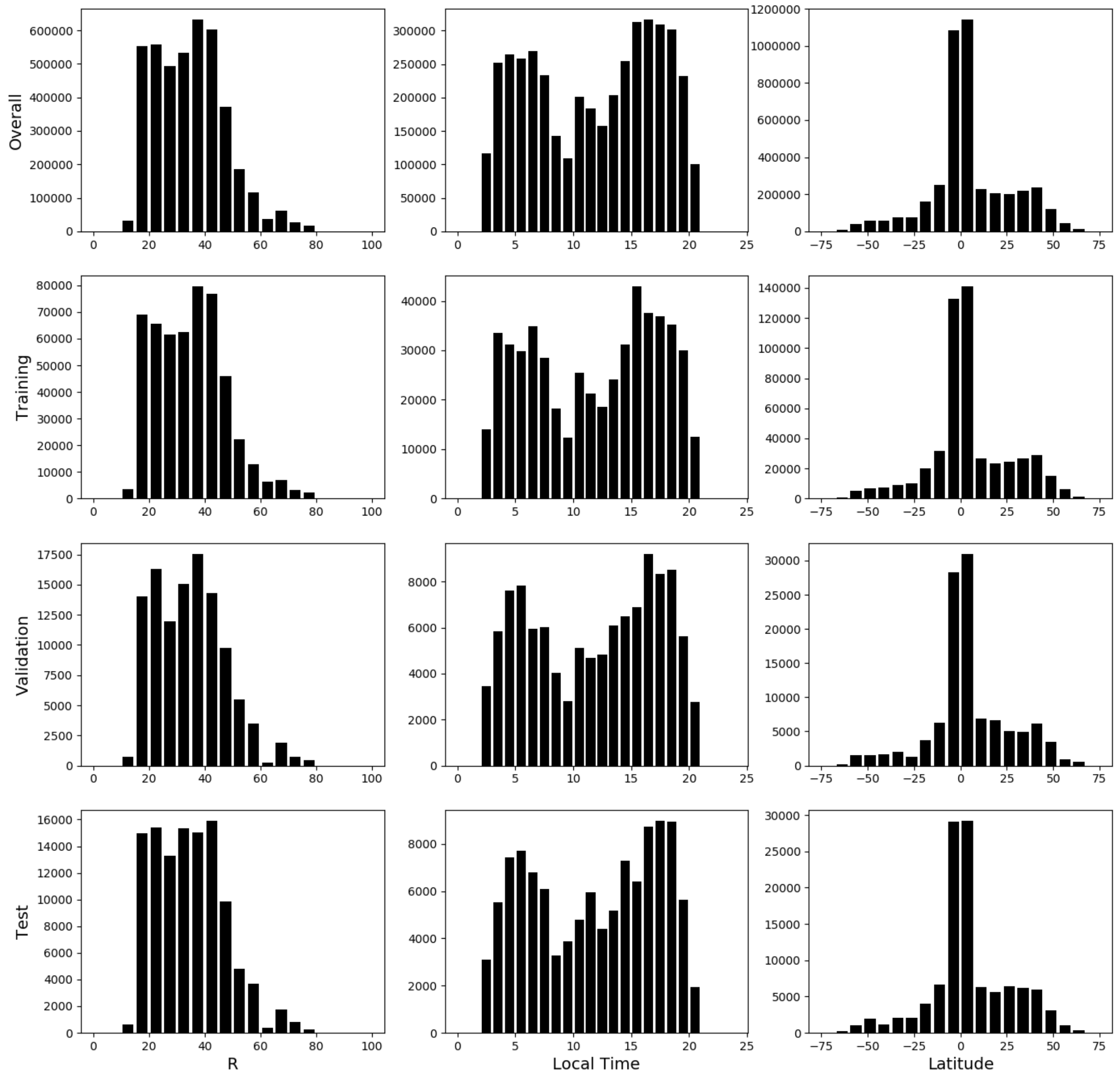
Large year-to-year variations in the orbit for Cassini means time-based splitting on a yearly basis (i.e., reserving an entire year for training) will result in a biased model

Instead, utilize time-based splitting on a smaller scale, with dedicated ”buffer” regions between the training, validation and test sets that are discarded

- Ensures no overlap between the three sets
- Should have relatively consistent S/C locations, preventing a biased model



- 105 hours for training (70%)
- 22.5 hours for test/validation (15%)
- 6 hour buffer periods between data (18 hours total)



Seeking training, validation and tests sets to be as evenly distributed in S/C location as possible

Feature Importance and Data Availability

- Interested in the impact of feature selection on the performance of the classification algorithm
- Features included:
  - 1-minute interpolated magnetometer (MAG) data in KRTP coordinate frame: |B|, B_Θ, B_R, B_Φ
  - 10-minute interpolated MAG data
  - 10-minute interpolated CAPS/CHEMS/LEMMS data
    - Explored using the full dataset (194 features)
    - Explored using a subset of the dataset deemed most important during manual boundary selection by scientists
      - CAPS/ELS 10eV electrons
      - CAPS/ELS 100 eV electrons
      - CAPS/ELS 10 keV electrons
      - CHEMS 4 keV protons
      - CHEMS 7 keV protons
      - LEMMS 40 keV protons
- For RNN, needed large quantities of data so only explored the MAG data
- For other classification algorithms, explored using various combinations
- **Algorithms are given zero knowledge about the S/C location**
  - Location information is used to ensure no bias is present in the training, validation or test sets
  - Location information used in interpretation of the results

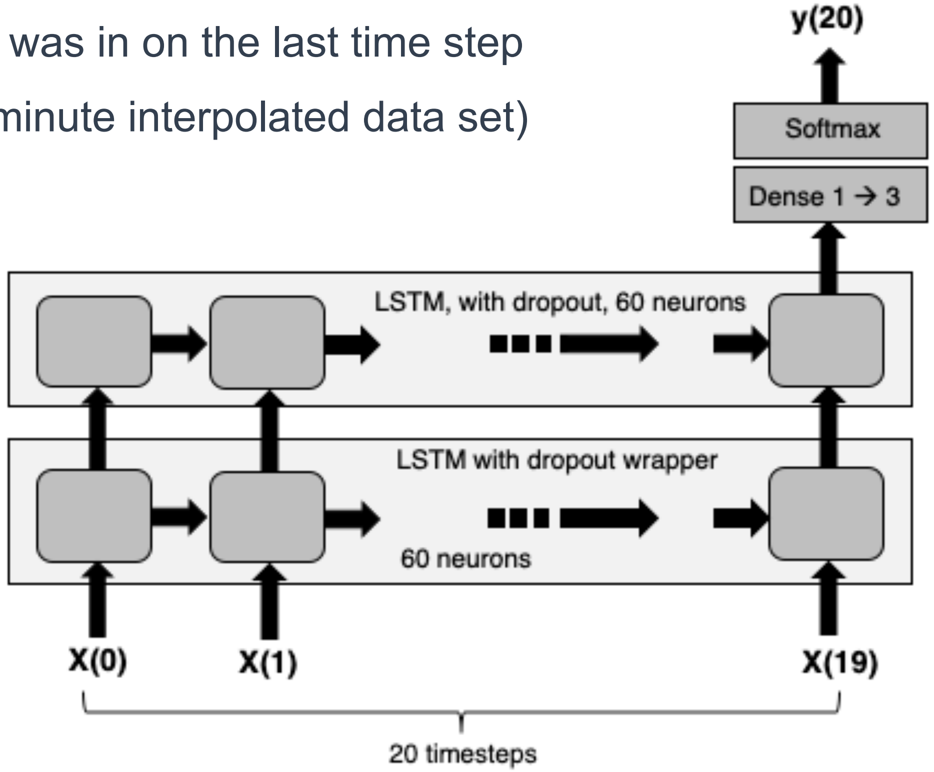
Classification Algorithm Approach

Simple classifier approach: Classify the region the spacecraft was in using only a single time step of data

- Algorithms explored: **Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR)**
  - Tuning for the RF approach included the number of trees and the minimum number of samples needed to split a branch
  - Explored the impact of different data sets on the resulting algorithm accuracy (MIMI – CHEMS & LEMMS, MAG, and CAPS) – using 10-minute interpolated data sets

RNN approach: With time series as input, classify the region the spacecraft was in on the last time step

- Due to data availability, were limited to only using magnetometer data (1-minute interpolated data set)
- Time series may or may not include a boundary crossing
- Iterated on:
  - Number of layers of RNN LSTM (1 or 2)
  - Number of neurons within the LSTM layer
  - Length of the time series (20, vs. 40 vs. 60 minute iterations)
- Controlled for overfitting by:
  - Including dropout
  - Early stopping on training when validation loss plateaued

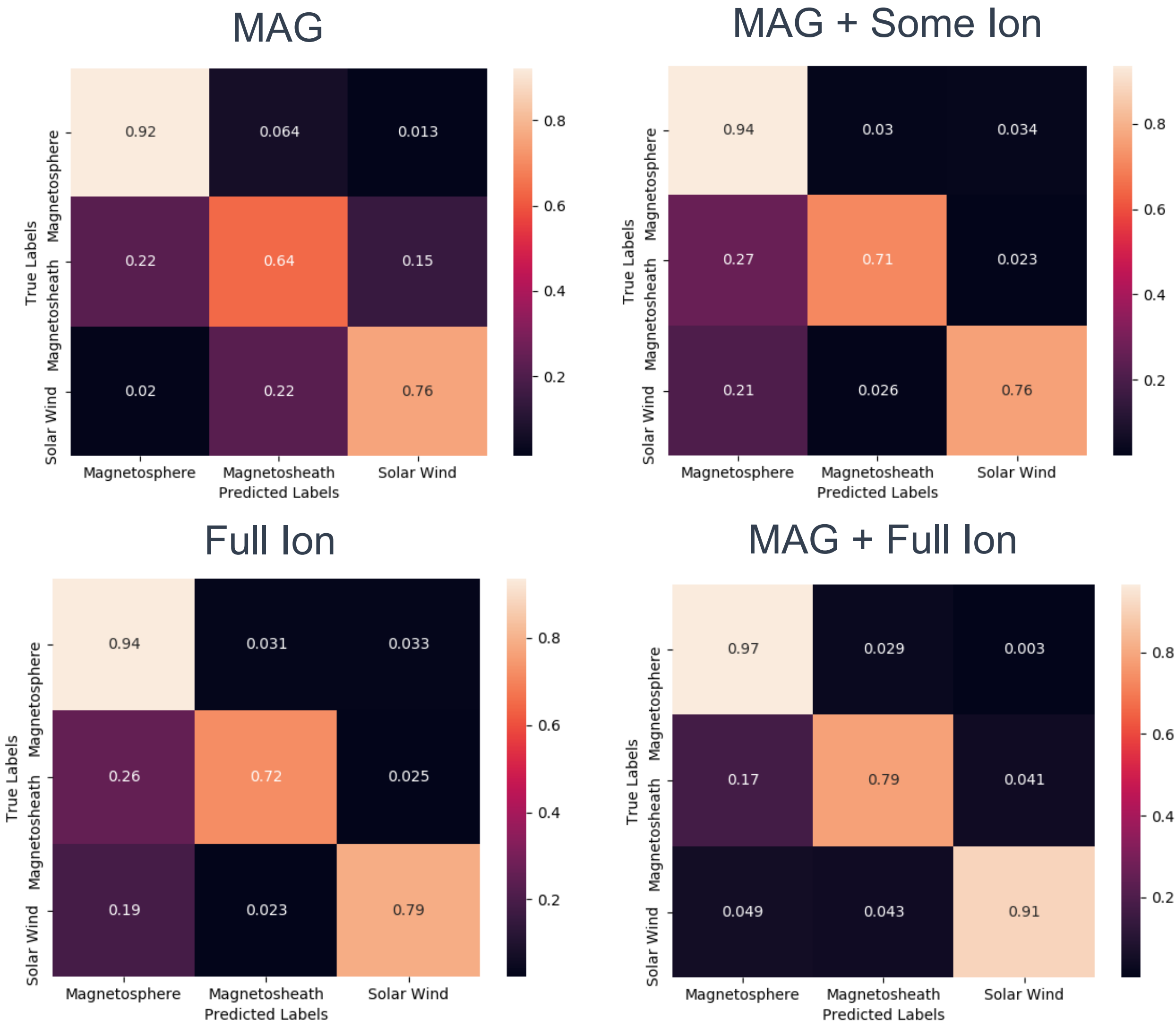


RF/SVM/LR Classifier Results

Dataset	Classifier	Num. Features	Train Accuracy	Test Accuracy
MAG	SVM	4	0.785	0.785
MAG	LR	4	0.788	0.788
MAG	RF	4	0.839	0.822
Some Ion	RF	6	0.728	0.740
MAG + Some Ion	RF	10	0.887	0.871
Full Ion	RF	194	0.875	0.861
MAG + Full Ion	**RF**	**198**	**0.957**	**0.913**

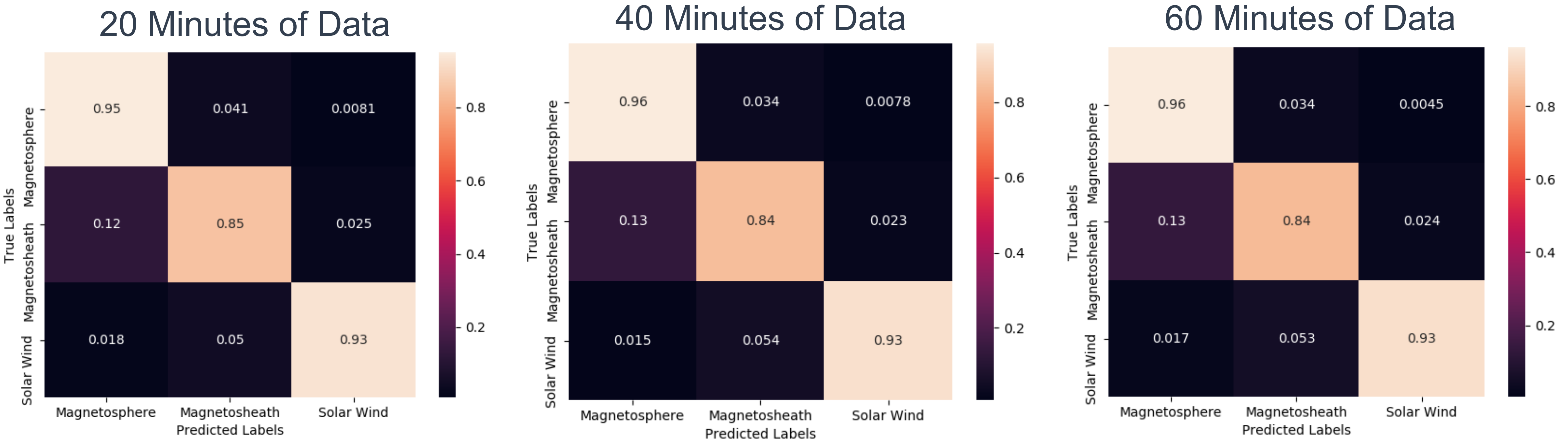
- Random Forest approach substantially out-performs logistic regression and SVM approaches for every feature combination
- Best combination of features is magnetometer dataset plus full MIMI and CAPS datasets
    - MAG data alone does fairly good job at discriminating between the different regions; confusion stemming around boundary transitions
    - Adding in the full set of ion data strongly increases the performance on the magnetosphere and solar wind regions, with some confusion still around the magnetopause

Confusion Matrices for RF Models



RNN Results: Predictions in the Absence of a Boundary Crossing

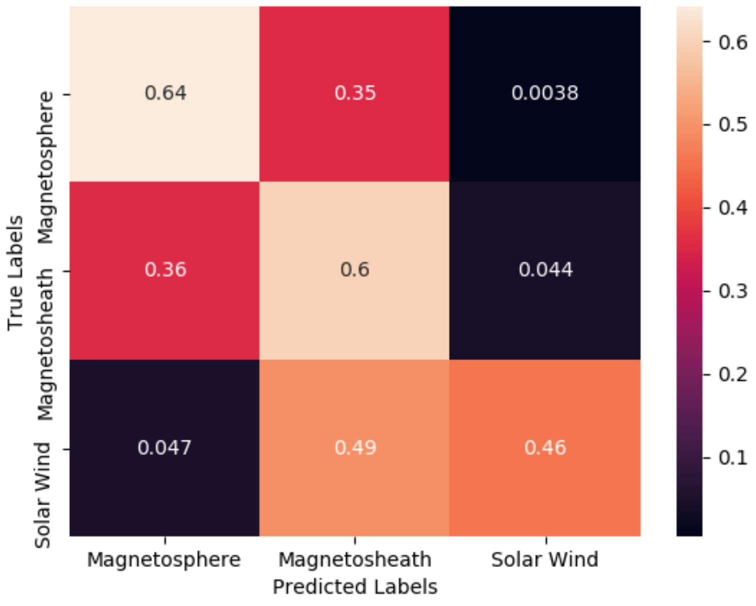
Maximum performance achieved ~ 92.5%



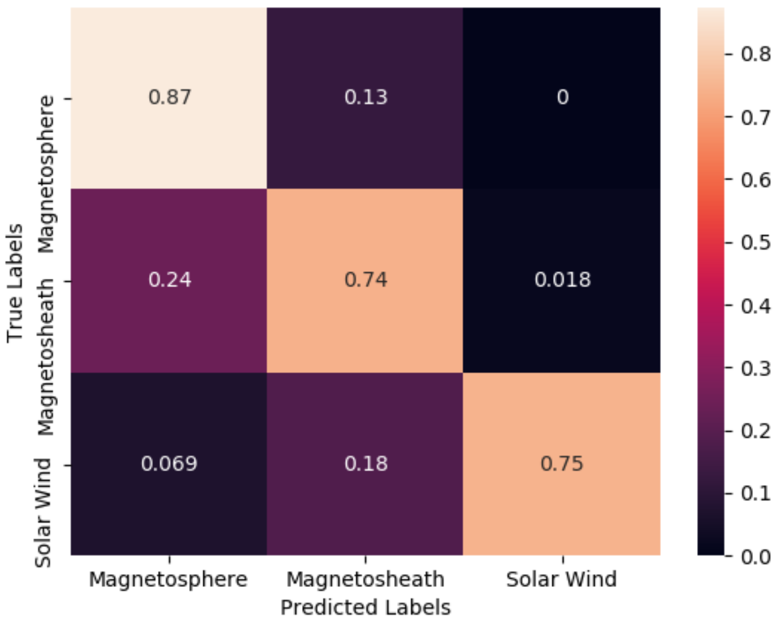
- Find increasing accuracy as move to deeper and larger networks, but also have increasing likelihood of overfitting
- No significant differences between the various RNN models in classifying data segments which do not contain a boundary
    - Strongest performance when classifying Magnetosphere or Solar Wind
    - Weakest performance in classifying Magnetosheath mainly due to confusion with the Magnetosphere

RNN Results: Predictions in the Presence of a Boundary

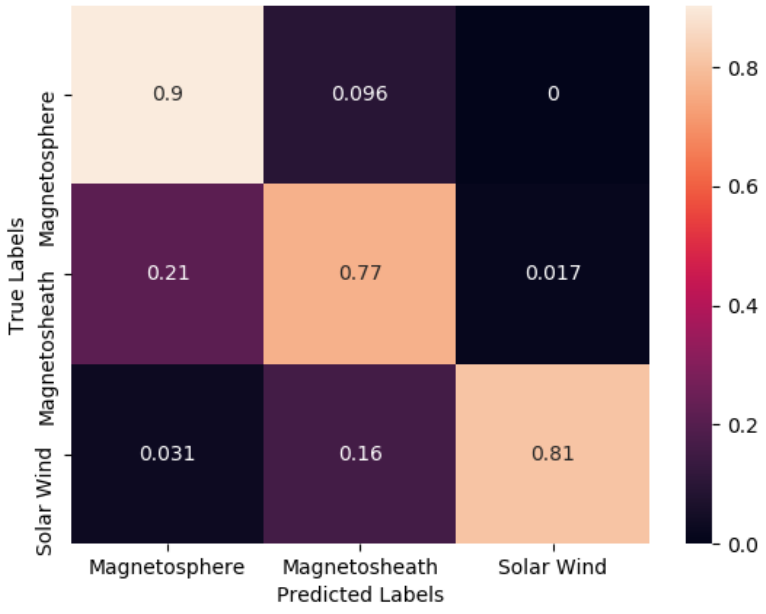
20 Minutes of Data



40 Minutes of Data



60 Minutes of Data



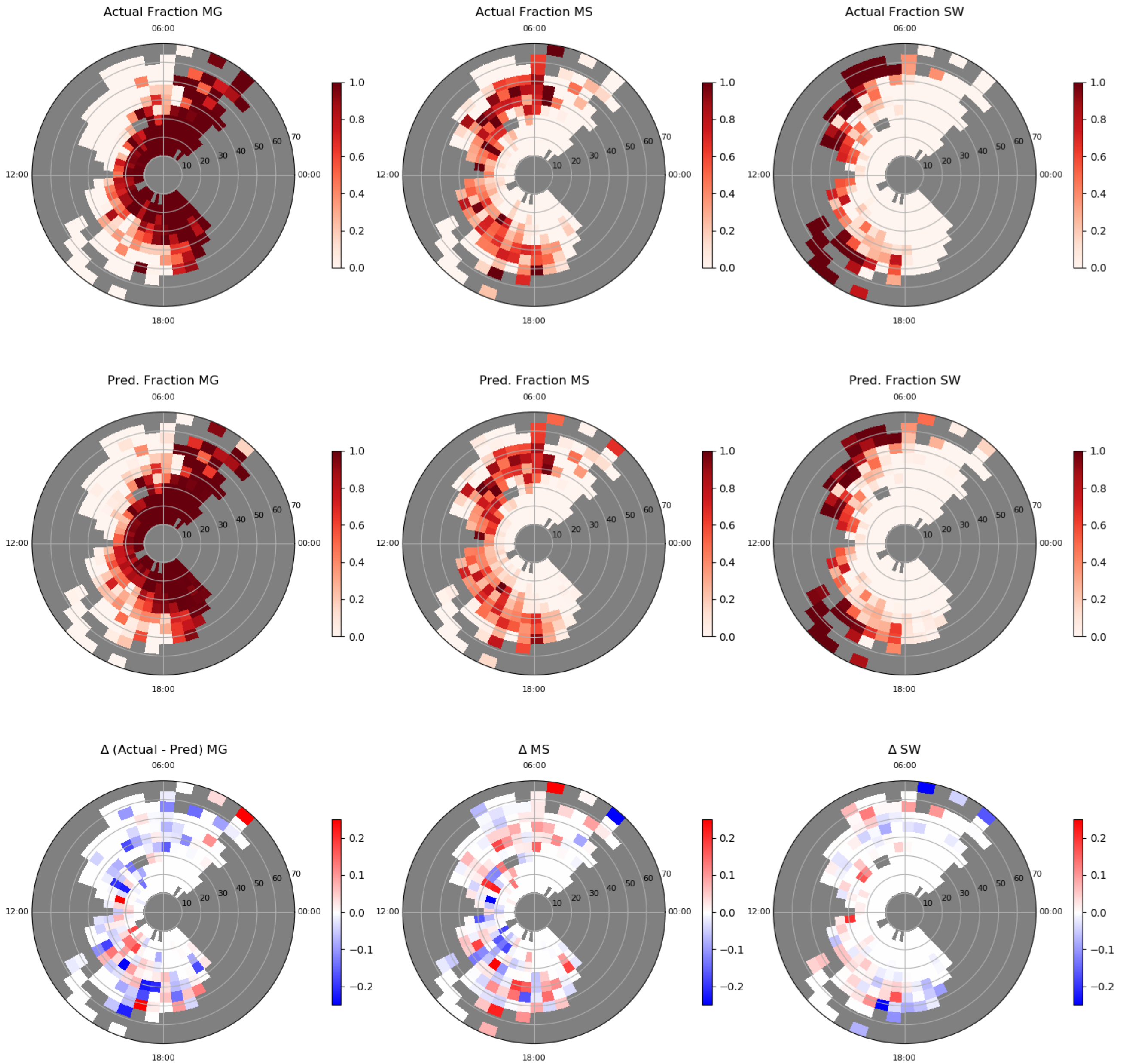
For samples which contain a boundary, much better predictions of the region after the transition as we move to longer time series

- Largest jump in improvement appears as transition from 20 minutes of data fed into RNN to 40 minutes of data
- Modest improvement, but significantly larger network required for 60 minutes of data

Hypothesize that improvements in accuracy are due to having better understanding of the gradients in the feature vectors → gradients are more significant than feature values for classifying a time segment

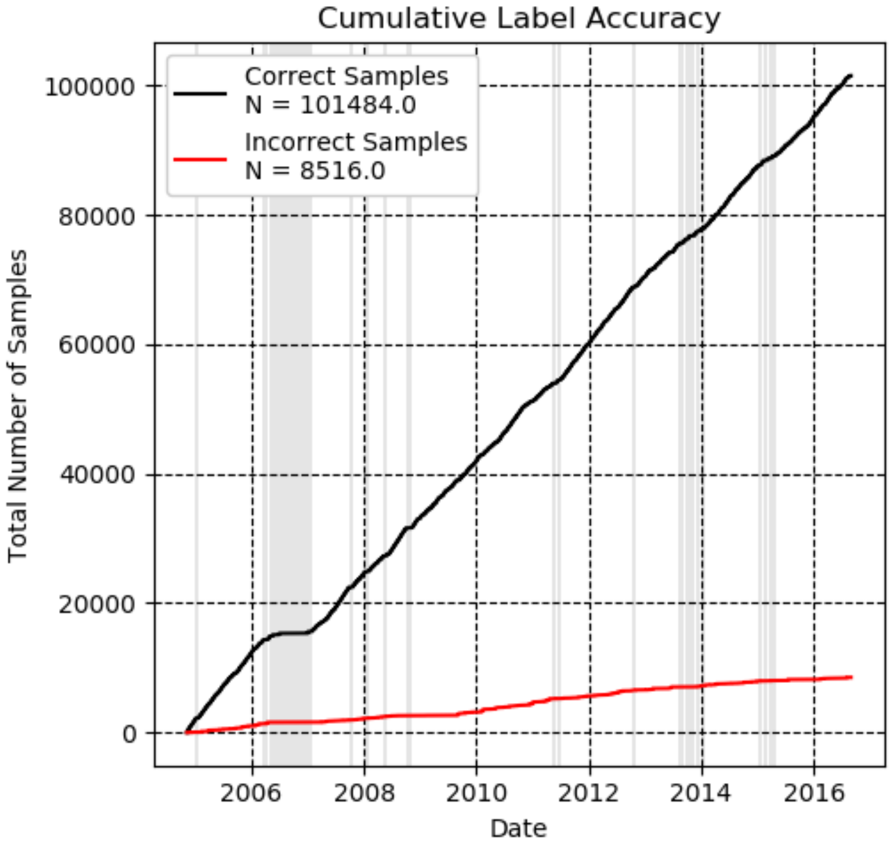
RNN Results: Understanding Where Errors Occur Spatially

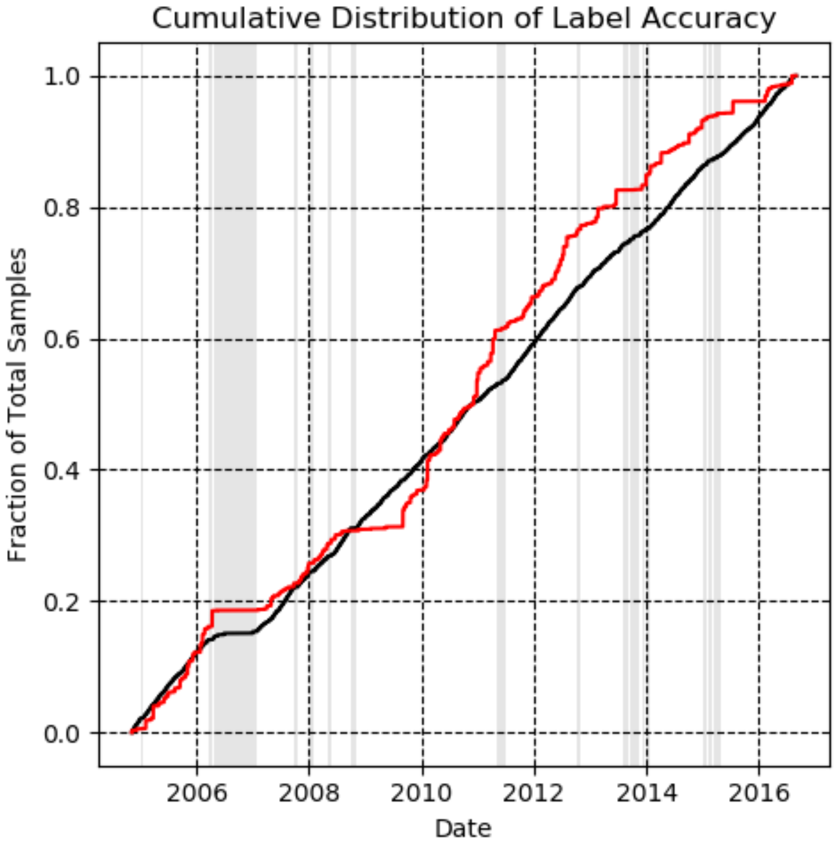
- Results shown for 40-time-step, 1-layer model
- No information on the spacecraft's location was given to the model, but results show the physicality of the problem was learned
  - Magnetosphere is closest to the planet, followed by magnetosheath and solar wind
  - |B| provides a clear indication of the distance to the planet (higher as you move closer in to Saturn)
- Areas of confusion appear to be near the boundary crossings
  - Over-prediction of magnetosphere appears to coincide spatially with underprediction of the magnetosheath
  - Solar wind is likewise confused with the magnetosheath
  - Need to investigate outlier "bins" where there was substantial over-/under-prediction (blue/red bins)
    - Possible bias in spacecraft latitude in these areas?



RNN Results: Understanding Where Errors Occur Temporally

- Bias in how Cassini's orbits were planned leads to discrepancies in where the errors occur based on time
- Are there particular S/C locations where we are more likely to get a prediction wrong?
- Are there abnormal feature values occurring in areas that are predicted incorrectly?

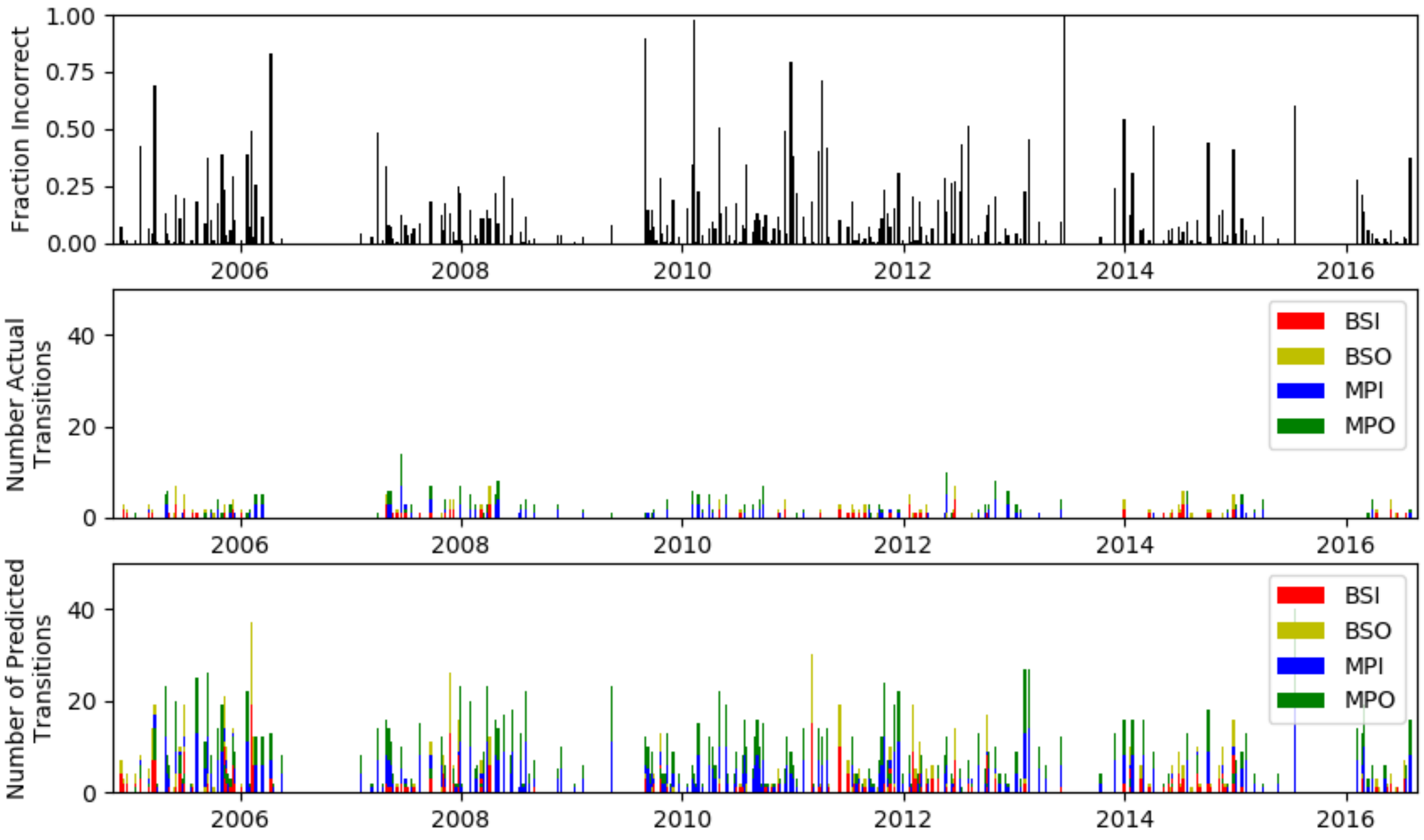




The correctly classified samples (black line) appear to accumulate at a constant rate. The incorrectly classified samples (red line) appear to have large chunks of accumulation, showing that there are particular orbit locations where the model fails consistently.

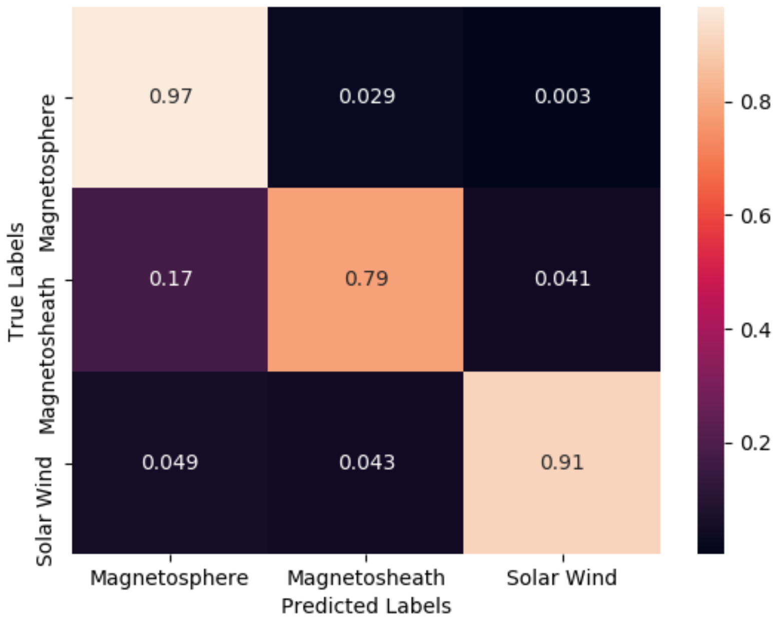
RNN Results: Understanding Where Errors Occur Temporally

- Look at each testing interval since the test samples are separated by design on a weekly basis
  - Clearly have large spikes in errors for particular intervals (also indicated by the CDF)
- Difference in the number of boundary crossings in a test interval
  - Are these real small-scale boundary transitions?
  - Alternatively, is the model unstable?
- Overall the BSI/BSO and MPI/MPO crossings that are predicted appear to coincide with those labeled

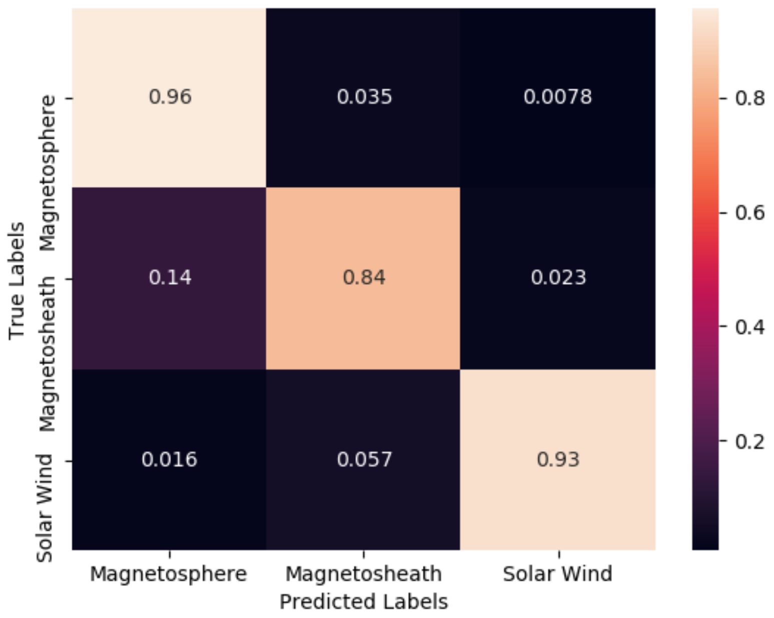


Comparison between RF with full ion + MAG dataset and RNN with just MAG data

RF with MAG & MIMI & CAPS



RNN with only MAG



Best RF Result: 91.3%
Best RNN Result: 92.5%

RF approach uses significantly less data (10 minute resolution versus 1 minute resolution of RNN), but each data point is much richer in feature depth (~200 features versus 4 – 8 for RNN)

RF is only predicting on a single time step, while RNN uses 20 – 60 time steps

RF is able to approach the accuracy of the RNN, only suffering with the magnetosheath predictions

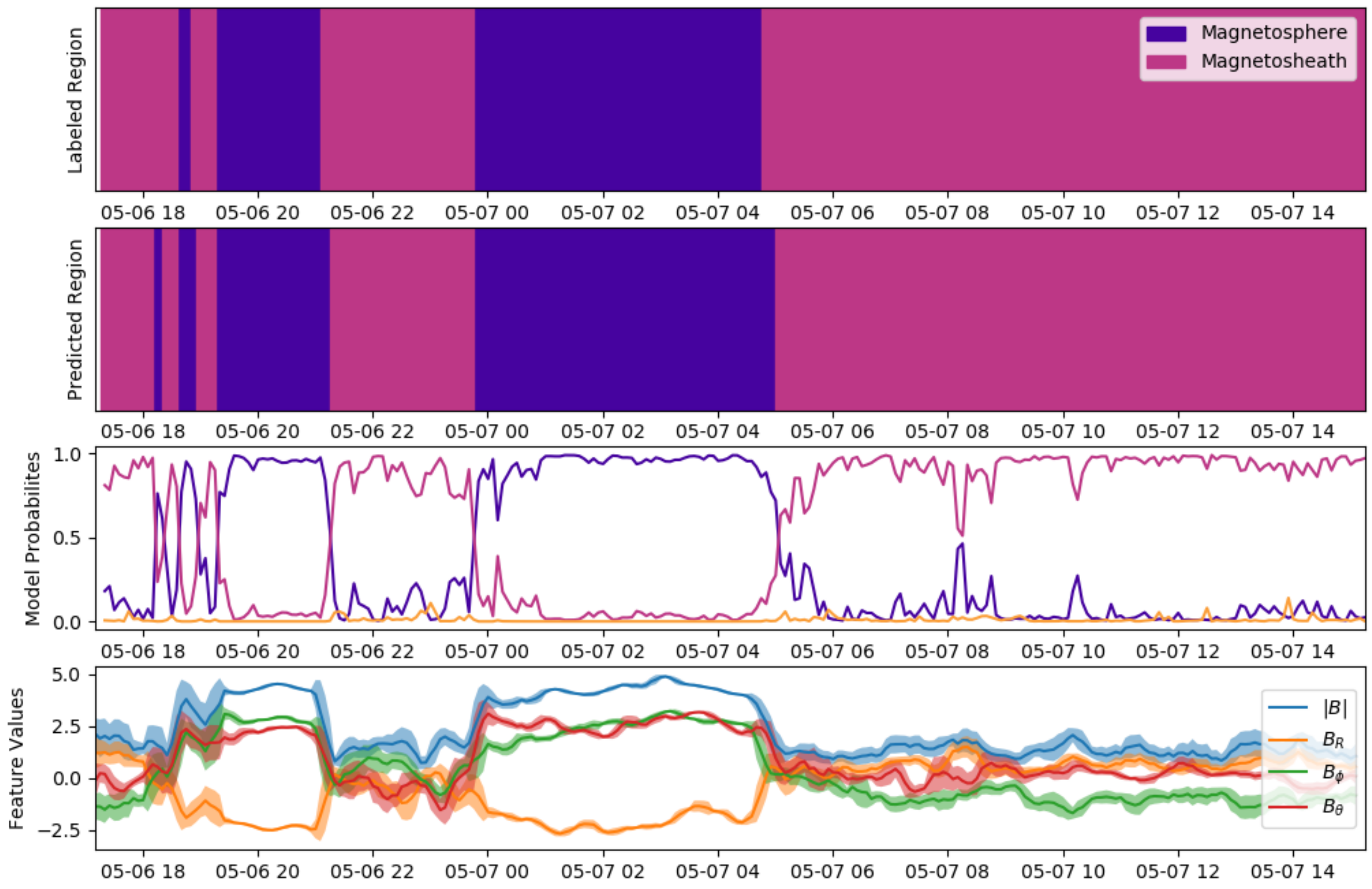
RF could likely be run on current spacecraft hardware

Conclusions

- Mission design significantly biases the dataset
  - Despite 12 years of relatively continuous data collection, Cassini only sparsely sampled the entire magnetosphere and magnetosheath around Saturn
  - Need to adequately capture the diversity of the dataset in the training, validation and test sets
  - Incremental parsing of training, validation and test sets with buffer periods insures each set is unique while representative of the entire orbit
- Using only magnetometer data can provide relatively accurate classifications of different regions when used with a sufficiently complex model
  - Maximum RNN accuracy achieved is ~92% on unseen test set
- Much simpler models, given more feature-rich data can perform nearly as well
  - Maximum RF accuracy achieved is ~91% on unseen test set
  - Substantially less data is needed to train
- Simpler models may be feasibly run on-board spacecraft with current hardware
  - RAD-hard GPUs not yet commercially available
  - Simpler ML models do not require fancy hardware

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

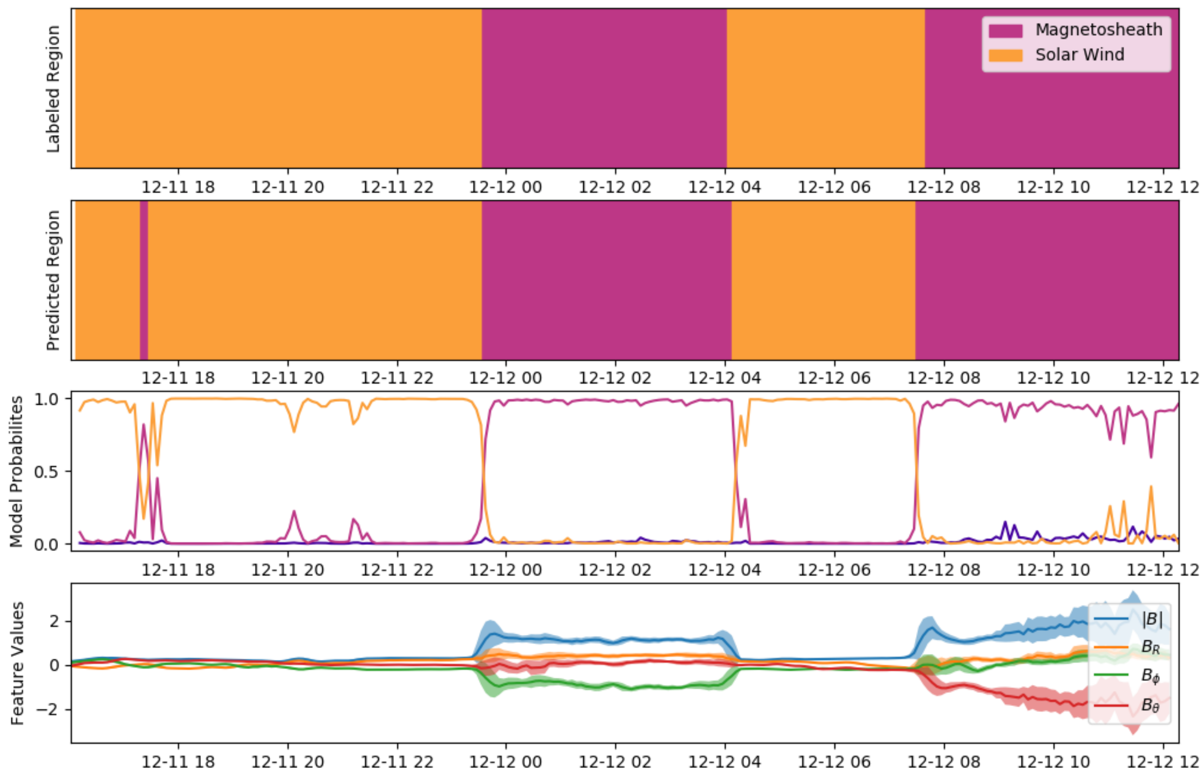
Understanding the Results in Time

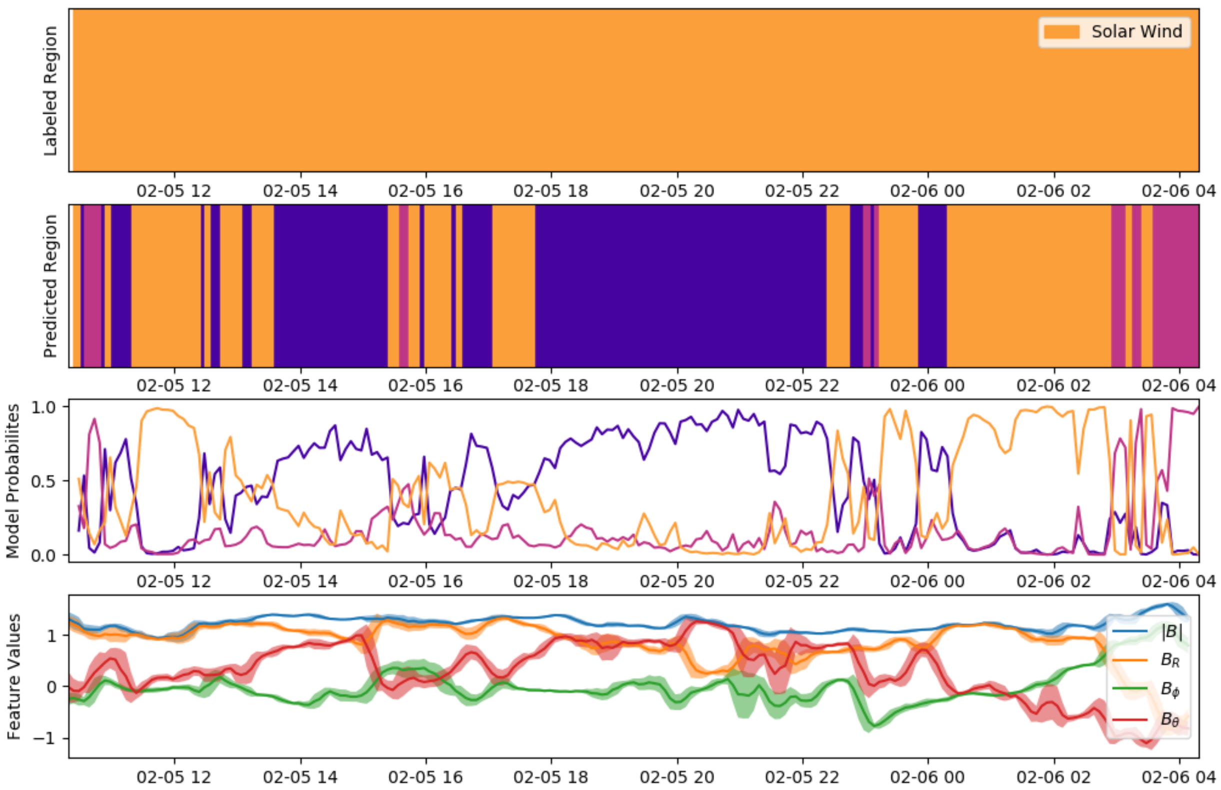


Is the RNN model generally picking up on boundary transitions?

- Where is it accurately predicting the boundary? Where is it getting it wrong?
- How is the S/C approaching the boundary? S/C speed/angle of attack relative to the boundary movement?
- Is it picking up on finer-scale boundary processes that are real? Or is the model overtrained?
- How does the model change when using longer or shorter time frames to classify the end point?
  - Hypothesize that longer time frames will allow for better classification (model has more context)

Clean Transitions versus Messy Transitions





Variance of the features appears to be more important than the absolute values for classifying a particular sample → longer time sequences should provide more context for the variance of a particular sample and provide better classifications

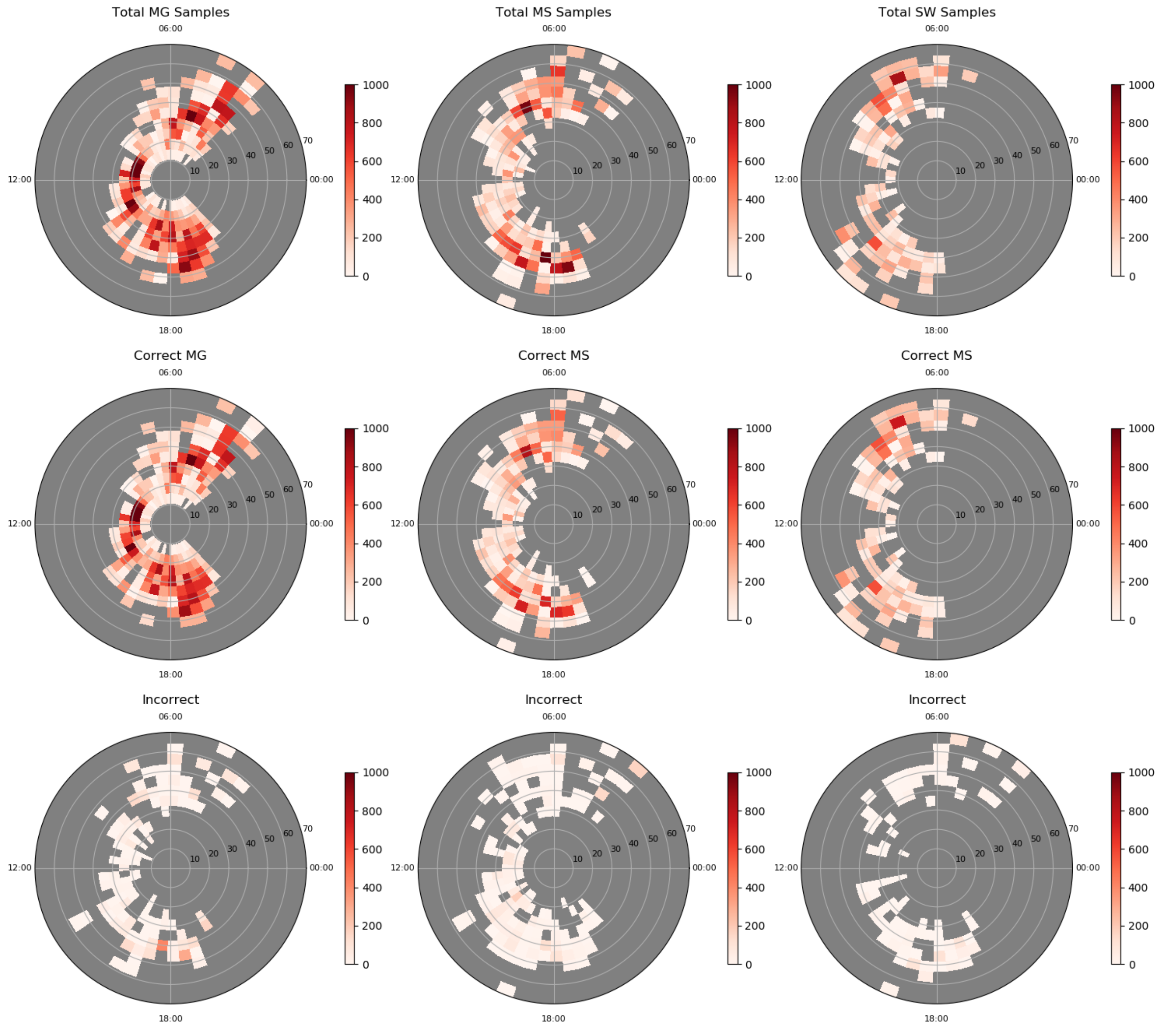
RNN Results

Model Type	Train. Set Accuracy	Val. Set Accuracy	Test Set Accuracy	Train. Set Loss	Val. Set Loss	Test Set Loss
RNN, 1-layer, 20 step	0.9286	0.9244	**0.9174**	0.1965	0.2313	**0.2361**
RNN, 2-layer, 20 step	0.9269	0.9218	**0.9143**	0.1974	0.2354	**0.2395**
RNN, 1-layer, 40 step	0.9400	0.9267	**0.9226**	0.1655	0.2234	**0.2222**
RNN, 2-layer, 40 step	0.9428	0.9271	**0.9241**	0.1565	0.2242	**0.2213**
RNN, 1-layer, 60 step	0.9435	0.9317	**0.9247**	0.1553	0.2151	**0.2161**
RNN, 2-layer, 60 step	0.9455	0.9267	**0.9220**	0.1474	0.2239	**0.2230**

Increasing accuracy but also increasing likelihood of overfitting as move to deeper networks and longer time segments. Over-fitting controlled by dropout and early stopping.

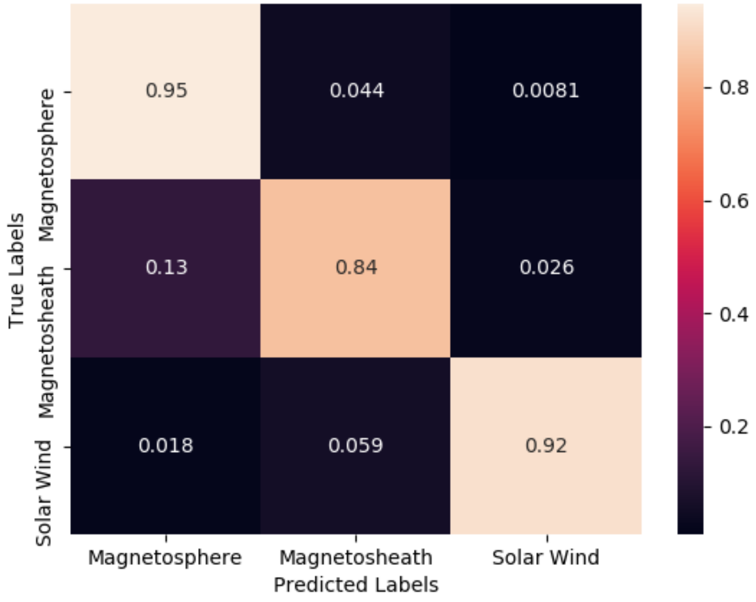
RNN Results: Understanding Where Errors Occur Spatially

- Results shown for 40-time-step, 1-layer model
- Clear bias in the sampling due to Cassini's orbits
- Limits imposed by the data processing also resulted in no predictions on the backside of the planet or within a close radius
- Model is able to correctly predict the regions for a vast majority of the samples

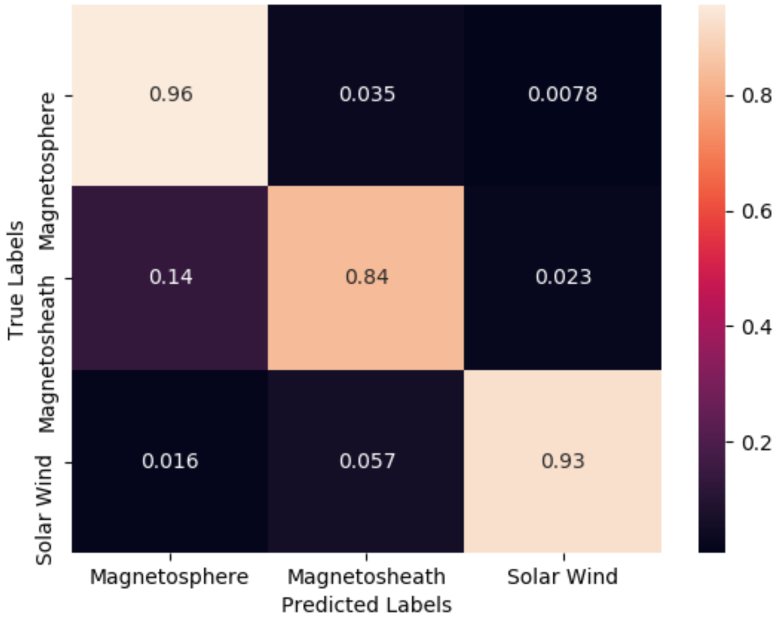


Overall Prediction Accuracy

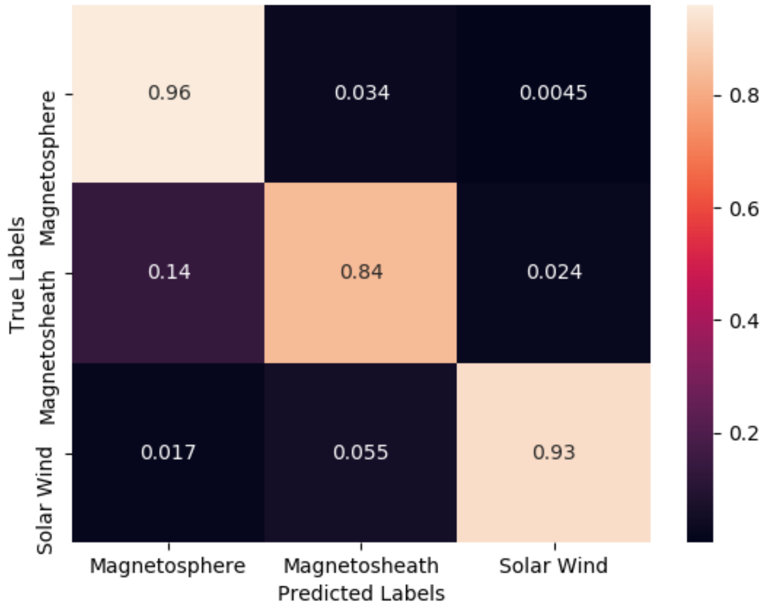
20 Minutes of Data



40 Minutes of Data



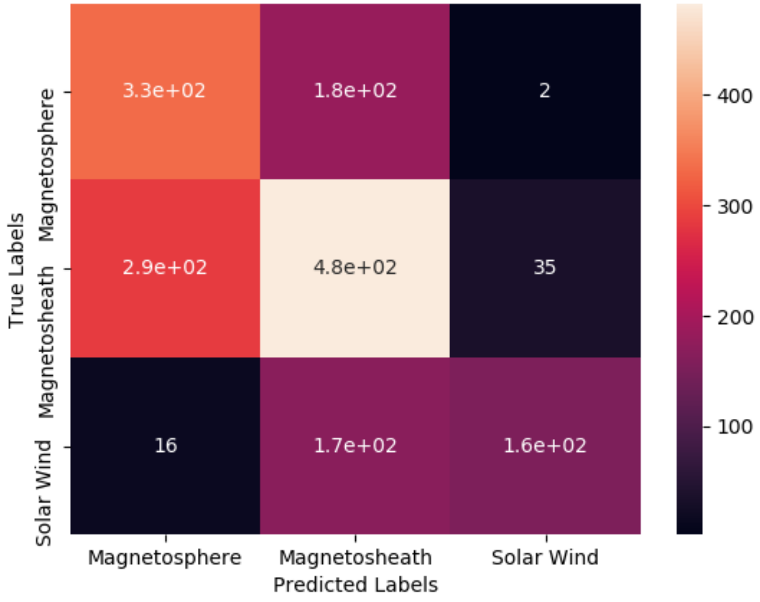
60 Minutes of Data



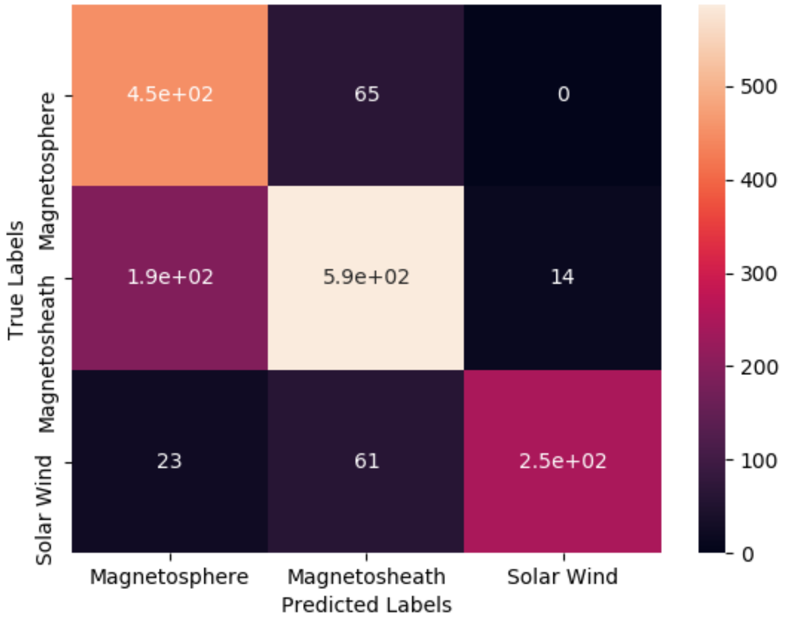
No significant difference in the overall prediction accuracy between the various models

Predictions in the Presence of a Boundary Crossing

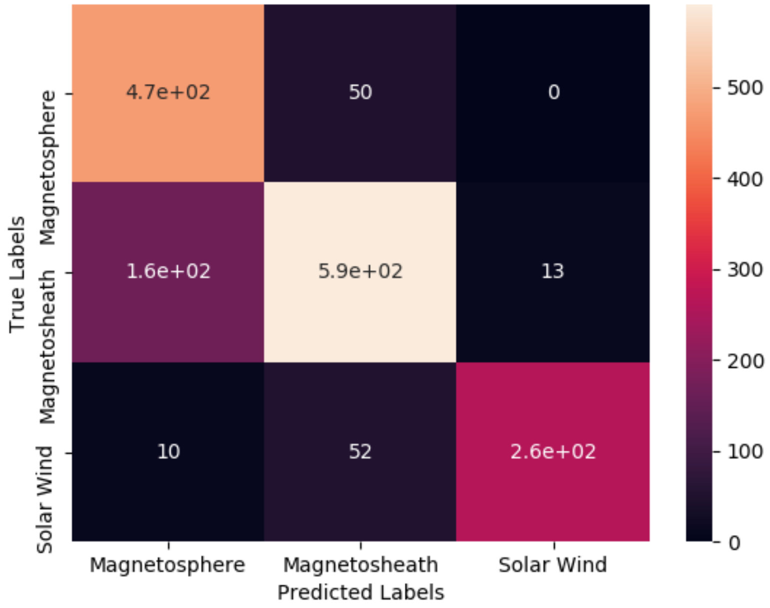
20 Minutes of Data



40 Minutes of Data

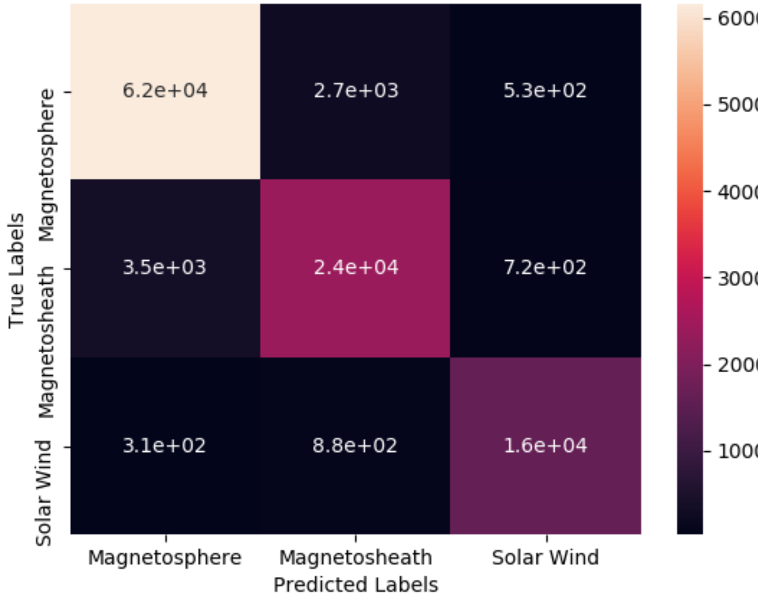


60 Minutes of Data

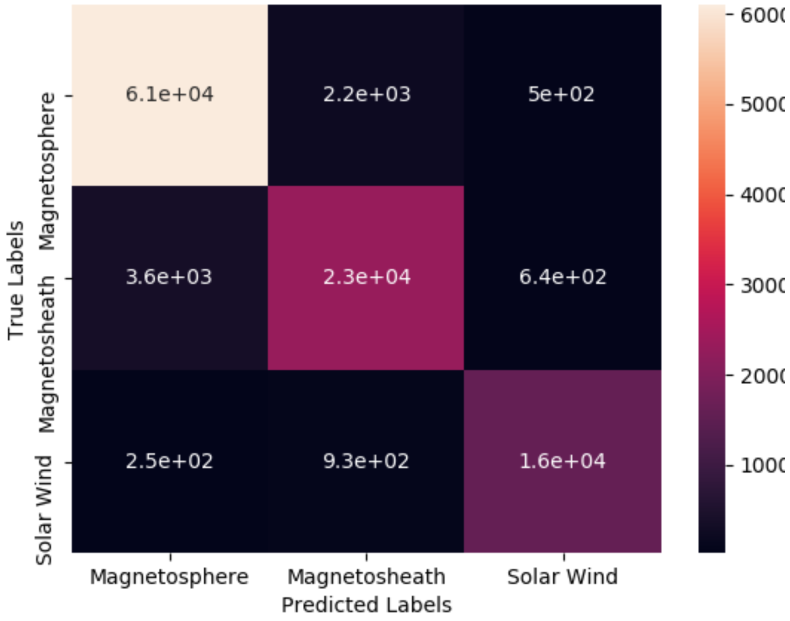


Predictions in the Absence of a Boundary Crossing

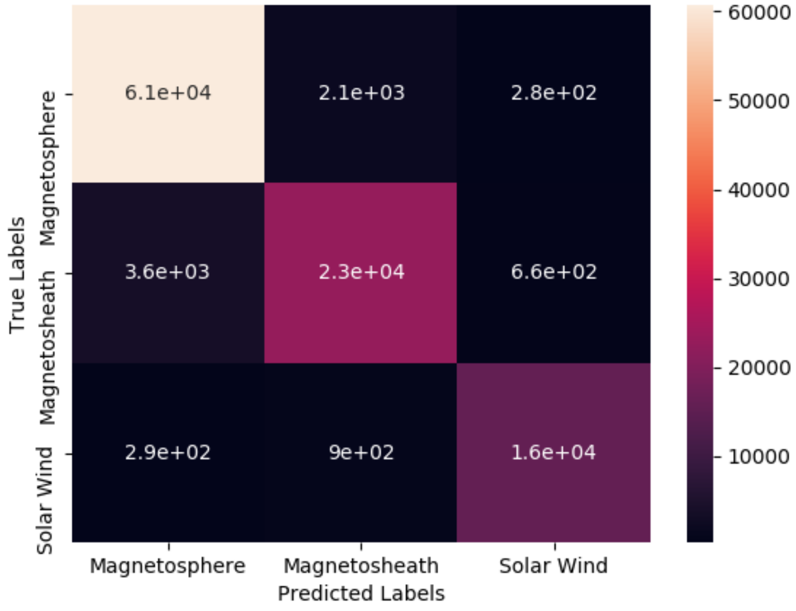
20 Minutes of Data



40 Minutes of Data

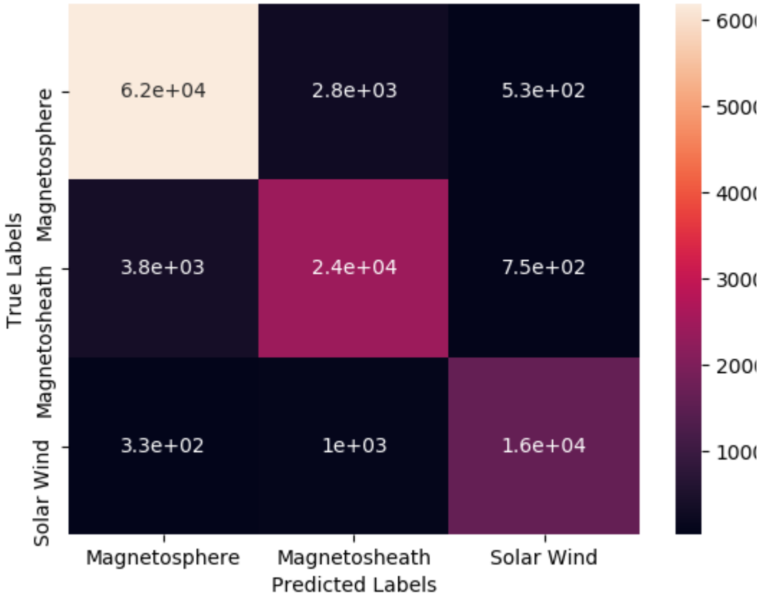


60 Minutes of Data

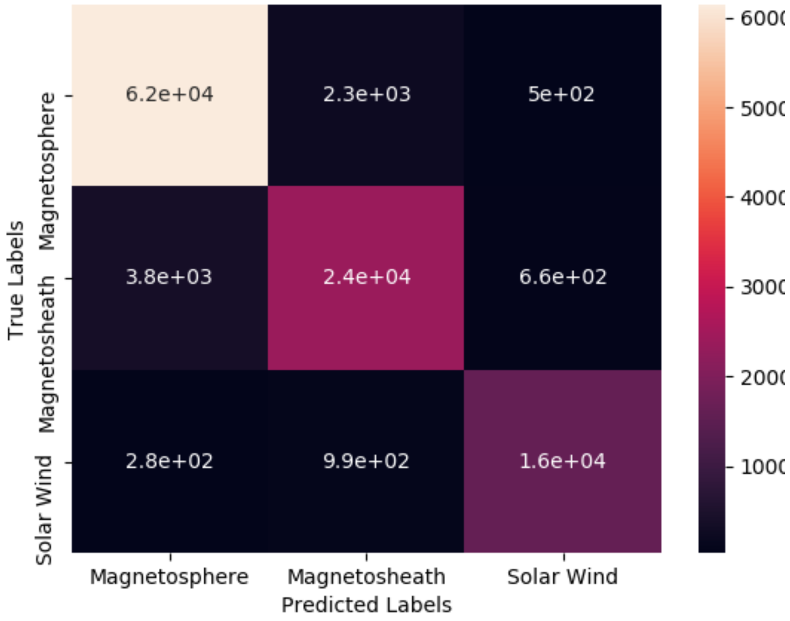


Overall Prediction Accuracy

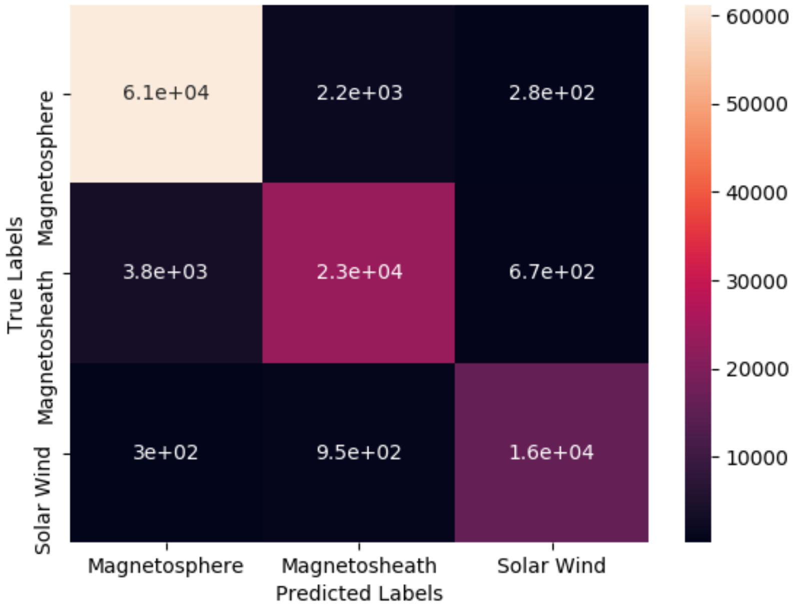
20 Minutes of Data



40 Minutes of Data



60 Minutes of Data



No significant difference in the overall prediction accuracy between the various models